DOCUMENT RESUME

ED 086 730 TM 003 381

TITLE	Symposium on an Alternative to Traditional Test Reporting.
SPONS AGENCY	Iowa Testing Programs, Iowa City.
PUB DATE	10 Nov 72
NOTE	55p.; Paper presented at Mid-South Research Association, November 10, 1972
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement Tests; *Computer Programs; *Educational Testing; Feedback; Parents; *Reports; Scoring; Standardized Tests; Student Evaluation; Student Testing; Teachers; *Test Results; *Verbal Communication
IDENTIFIERS	Madison Project; *Narrative Test Reporting; Test Reporting

ABSTRACT

This symposium presents an alternative to the numerical mode of reporting, and describes the work that has been done, and the work that is continuing, in the area of computer-generated narrative-format testing reports. Computer narrative testing reports are in a verbal form that eliminates the use of almost all numbers and technical terms. These reports have been designed for the pupil, his parents, and his teacher. The need for alternative reporting systems is discussed in terms of serving the unique needs of the examinee and others who receive reports of his test scores. The application of computer technology to test reporting covers a brief overview of verbal format testing reports and describes the Madison Project with its accomplishments and its limitations. The Iowa Project is discussed in terms of its objectives and the kind of narrative reports that are planned. Each of the audiences of the reports is discussed in terms of their needs and the kinds of reports that are available to them. The presentation, The Future of Narrative Test Reports, sketches the current plans for the Iowa Project and some long-range possibilities and applications for narrative format reports. (END)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 086730

TM 003 881

Symposium
on

AN ALTERNATIVE TO TRADITIONAL TEST REPORTING

Mid-South Educational Research Association
November 10, 1972

Support for planning the Iowa phase of this project, and for the development of this report, was made possible by a grant from the Iowa Testing Programs, University of Iowa.

CONTENTS

I. THE NEED FOR ALTERNATIVE REPORTING SYSTEMS 1
Marcia A. Whitney

II. THE APPLICATION OF COMPUTER TECHNOLOGY TO TEST REPORTING 5
Walter M. Mathews

III. THE IOWA PROJECT: IOWA TESTS OF BASIC SKILLS NARRATIVE REPORTS FOR PUPILS 18
Ralph Balfour

IV. NARRATIVE TEST REPORTS FOR PARENTS 30
Earl Stephan

V. NARRATIVE TEST REPORTS FOR TEACHERS 37
Stephanie D. Alnot

VI. THE FUTURE OF NARRATIVE TEST REPORTS 45
Thomas B. Taft, Jr.

I

THE NEED FOR ALTERNATIVE REPORTING SYSTEMS

Marcia A. Whitney

For years, a common feature of public schools has been the acquisition of readin', 'ritin', and 'rithmetic skills. It would seem that today we could add a fourth commonality to the three R's: testin'. Goslin estimates (1967, p. 1-2) that each of the fifty million or more school children in the United States takes, on the average, three standardized tests each year. Testing has become a big business, and test-builders, publishers, and users have each examined and made recommendations regarding test use and abuse.

Given that standardized tests are not perfect measuring instruments, they do provide a source of information about the pupil that is not subject to individual teacher idiosyncracies and individual school curriculum offerings (Ebel, 1971). Obtaining the information, however, is only one step in the testing process. An equally important step is providing feedback to the appropriate persons about the resultant performance of the pupil. But who are the appropriate persons? And how can the results be most meaningfully and responsibly reported?

Ebel (1964) maintains that test-makers have been too reluctant to share information about test results with the public. If one is to begin making informed choices about his future, he needs as much information as is available to him about his own abilities. Certainly, then, one of the persons who should be informed about test performance is the person who took the test. As the pupil matures and becomes more autonomous in educational decision-making, such information becomes increasingly important. It is important even to the elementary school child in that it begins to establish that achievement tests are a source of information about oneself, and not just a yearly respite from the daily activities in the classroom. Goslin indicates (1967, p. 24) that many believe that it is precisely during this period of relative malleability of self-conception that more systematic information about the child's abilities should be available to the child. The teacher and the parents assist in educational decision-making for the elementary pupil, and share in aspirations, so they, too, need to be informed about the pupil's performance on standardized tests.

These persons--the pupil, the parents, the teacher--each make some unique use of the test performance results. The pupil's needs are perhaps the most personal: it is his/her performance that is being reported, whose skills are being evaluated. The pupil needs to get the information in language that is understandable, and in a manner that

is positive, not punitive. The parents share with their child a concern for his/her particular performance. Additionally, however, they can use the results to structure future, informal learning situations for the child. The teacher is concerned with the performance of each pupil in the classroom, and faces the challenge of providing classroom activities that meet the educational needs of all pupils. The teacher also must view the entire class in relationship to other similar classes, and evaluate the results of the school's particular curriculum outcomes against the more general outcomes reflected by the standardized achievement test.

All these persons share in their need for test reports that are clear, concise, and useful. They each need answers to the following questions:

1. How did the pupil do on this achievement test?

2. How much did the pupil grow since the last similar test?

3. Does the pupil have any <u>specific</u> strengths and/or weaknesses?

4. What can, or should, I do in response to this report?

Traditionally, the results of standardized achievement tests are returned to the school officials. Goslin reports (1967, p. 32) that elementary teachers receive score reports on pupil performance about 80% of the time and have free access to scores in the remainder of cases. Feedback is clearly available to the teachers, and they are in a position to serve as the disseminators of information about test performance to the pupil and the pupil's parents. Yet Goslin's data (1967) reveals that teachers do not routinely transmit this information: feedback of standardized test results to parents and pupils in general occurs less than 8% of the time.

This low frequency is in large part accounted for by widespread reluctance to disseminate intelligence test scores. But while the data on dissemination of achievement test results is somewhat more encouraging, there is evidence that there is much room for improvement in both frequency and quality of communication. Goslin advances several apparent reasons for the inadequacy of communication about test results. There are many demands placed on the classroom teacher's time, and reporting test score results to each pupil and accompanying parents places an additional burden. While achievement test publishers have taken pains in their accompanying test manuals to guide meaningful interpretation and use of test results, some background in measurement and evaluation on the part of the teacher would undoubtedly facilitate their use. Not all teachers have such background, and they may not feel comfortable attempting to interpret test score results to others. It would seem, therefore, that lack of time and training have resulted in limited application of materials provided by test publishers.

Other reasons that teachers fail to share information might be more affective than technical. Some teachers fear that they will be evaluated on the basis of these results, and that test scores will be used to judge the quality of their teaching. Some teachers believe that only a "professional" person should have access to test information, and guard against any leak of information.

There is clearly no single answer to the question: Why do so few teachers report test results to pupils and parents? There are various alternative solutions to the dilemma posed by the question. More training in measurement could be required of all teachers, so that score interpretation would be knowledgeable and accurate. Teachers could prepare carefully written reports for pupils and parents so that there would be no misunderstanding of the test results on the part of the latter persons. These alternatives would provide some resolution to the technical issues and probably the affective ones as well, but they exacerbate rather than alleviate the problem of insufficient time.

Fortunately, another alternative exists that provides a partial answer to both insufficient training and time: computer technology. Some beginnings have been made in computer-generated score reports on standardized tests. The project being reported in this symposium is an attempt to develop this alternative method of score reporting for the Iowa Tests of Basic Skills.

REFERENCES

Ebel, Robert L., "Standardized Achievement Tests: Uses and Limitations," The National Elementary Principal, (1961), 40, 29-32.

Ebel, Robert [illegible], "The Social Consequences of Educational Testing," School and Society, (November 14, 1964), 331-334.

Goslin, David A., The Search for Ability: Standardized Testing in Social Perspective. (New York: Russell Sage Foundation, 1963).

Goslin, David A., Teachers & Testing. (New York: Russell Sage Foundation, 1967).

II

THE APPLICATION OF COMPUTER TECHNOLOGY TO TEST REPORTING

Walter M. Mathews

Organizations providing scoring and interpretation services are increasing in number and importance, primarily because of the speed, low cost, and complexity of scoring programs possible due to the introduction of computers into the processing system. Speed and low cost of scoring increase the use of tests. The great capacity of computers makes it economically feasible to use much more complex scoring keys than was formerly possible, and to relate scores and individual responses to population characteristics of persons whose tests are being scored. The use of these machines, therefore, may increase the quality of tests and test manuals.

A greater and more immediate impact may come from more complex and sophisticated interpretation of test scores which such machines make possible. Studies show that few test takers get feedback on their testing, and very few understand the meaning of the information given them, since most of it is coded and summarized in ways which seem more sensitive to the test processing needs than to the test takers' or counselor's needs (Holmen and Docter, 1968).

Let's look at four major attempts to utilize the computer to generate verbal testing reports.

Teaching Information Processing System (TIPS)

Kelley (1968), a professor of economics at the University of Wisconsin, has developed a program, which he calls TIPS, to assist him in teaching. TIPS involves periodic collection of information from students regarding either their understanding of course materials or their reaction to various aspects of course presentations. TIPS provides a means of efficiently utilizing this information for instructional purposes. The information, which is collected on specialized forms suitable for machine processing, is composed of student responses to a series of multiple-choice questions. Surveys of six to twelve questions take about five to ten minutes to administer. Within a few hours this information is processed and summarized in three separate reports: one for each student, one for each section leader, and a third for the professor.

The student report contains a summary of his performance: his response to each question, the correct answers, and the total number of his correct answers. On the basis of this information, assignments for the forthcoming period are also indicated. The assignments (some required--some optional) vary considerably in nature, level, and intensity. A student scoring well may receive optional assignments and/or required work at a higher level. The student performing poorly may receive not only a heavy dose of required work but also a set of materials designed to bring him toward the mean class performance.

Additional information on the student report is generated on the basis of past as well as current performance. If the student has performed poorly over several surveys, he will be instructed on the student report to establish an appointment with the instructor or teaching assistant. If the student performed consistently and exceptionally well, he may be notified that a short paper may be substituted--at his option--for the midterm examination. A sample of a student report is presented as Figure II.1.

The teaching assistant report contains information to help him appraise the performance of his individual sections, including statistics on percentage correct by question or by concept, actual responses on the survey, lists of students required to establish appointments or tutorials, and so forth.

The professor's report is similar to that received by the teaching assistant, although the information available applies to all students enrolled in the course rather than only to the students in particular sections. With this information the professor may elect to alter lectures, section coverage, problem sets, or other teaching instruments for the forthcoming period. In summary, TIPS is a system for gathering and reporting objective and timely information useful for more effective teaching.

Figure II.1

SAMPLE COPY OF KELLEY'S TEACHING INFORMATION
PROCESSING SYSTEM (TIPS) STUDENT REPORT

TIPS
STUDENT PERFORMANCE SURVEY RESULTS
PRINCIPLES OF ECONOMICS (103)
PROFESSOR ALLEN C. KELLEY

TURNER, LAWRENCE
715648
SURVEY TAKEN 10/25/67
SECTION NUMBER AND TIME 2, 9:55 F
SECTION LEADER MISS GREEN

OUT OF A TOTAL OF 10 QUESTIONS, YOU CORRECTLY ANSWERED 3. THE QUESTION NUMBER, YOUR RESPONSE AND THE CORRECT ANSWER ARE PROVIDED IN THE TABLE BELOW. YOU ARE URGED TO MAKE SURE YOU UNDERSTAND THE NATURE OF ANY INCORRECT RESPONSES YOU MAY HAVE MADE.

SUMMARY OF SURVEY RESULTS

QUES. NUM.	YOUR ANSW.	CORR. ANSW.	QUES. NUM.	YOUR ANSW.	CORR. ANSW.
1	G	G	6	B	A
2	C	B	7	D	E
3	D	B	8	A	A
4	A	C	9	D	C
5	F	F	10	C	B

YOUR ASSIGNMENT FOR THE WEEK, TO BE HANDED IN DURING THE DISCUSSION SECTION ON 11/03/67, IS THE FOLLOWING --

PROBLEMS 1, 3 AND 4 ON HANDOUT 2C

ADDITIONALLY, YOU ARE REQUIRED TO WORK THROUGH CHAPTER 2 OF - MICRO-ECONOMICS, A PROGRAMMED BOOK, BY LUMSDEN, ATTIYEH, AND BACH. IT WOULD BE USEFUL TO CONSULT THE PROGRAMMED UNIT BEFORE YOU READ HANDOUT 2C.

TO TEST YOUR UNDERSTANDING OF THE MATERIALS, YOU MAY, AT YOUR OPTION, ELECT TO COMPLETE THE FOLLOWING PROBLEMS --

1. WORKBOOK, PP. 37-38.
2. HANDOUT 2A, PROBLEMS 3 AND 4.

NOTICE -- AN ALL UNIVERSITY LECTURE BY PROFESSOR MILTON FRIEDMAN, PAST PRESIDENT OF THE AMERICAN ECONOMIC ASSOCIATION, WILL BE HELD IN 6210 SOCIAL SCIENCE, 7:30 P.M., 11/01/67. THE TOPIC -- MONETARY AND FISCAL POLICY RECONSIDERED.

Preliminary Scholastic Aptitude Test (PSAT) Score Reports

The effective reporting of test results to admissions officers, guidance counselors, and the individuals who take educational tests is a rather more complex matter than might appear at first glance. It is apparent that reporting only numerical scores is hardly adequate. The statistical, psychological, and educational contexts which allow the user to infer relevant meanings must be provided as well. The problem is particularly acute for the programs in which the primary reporting target is the test taker himself.

Since the precise interpretation of mental test scores does require rather sophisticated insights into statistics, psychology, and education, it is common practice to have test scores reported to individual test takers by guidance counselors who are expected to have the necessary sophistication. Unfortunately, not all guidance counselors do have sufficient psychometric sophistication, nor do they have the time to prepare detailed analyses and to give individual interpretation of mental test results. Of course, most educational testing services prepare a wide range of interpretive materials for their testing programs to aid both the counselor and the student in interpreting test scores. But even with these aids, a fair degree of statistical sophistication is still needed for adequate understanding.

Helm and Harasymiw (1968) designed variable format computer-generated letters to be sent to examinees of the Preliminary Scholastic Aptitude Test as a report of their performance on the test. They found it necessary to prepare eighty distinct sentences where a sentence might have variable phrases imbedded. These eighty sentences were used to compose seventy-five distinct paragraphs which, in turn, were combined to produce the 100 letters needed to interpret all combinations of verbal and mathematical scores. They wrote a computer program to generate these letters and a sample of their report appears as Figure 11.2.

Figure II.2

SAMPLE COPY OF THE PRELIMINARY SCHOLASTIC
APTITUDE TEST (PSAT) SCORE REPORT

EDUCATIONAL TESTING SERVICE

PRINCETON, N.J. 08540

JANUARY 21, 1966

DEAR MR. LETR 99,

WE WANT TO REPORT TO YOU THE SCORES YOU EARNED ON THE PRELIMINARY SCHOLASTIC APTITUDE TEST YOU TOOK ON OCTOBER 9, 1965. YOUR APTITUDE FOR COLLEGE WORK IS OUTSTANDING. IF YOUR HIGH SCHOOL MARKS ARE CONSISTENT WITH THE HIGH SCORES YOU HAVE EARNED ON THE TEST YOU WILL HAVE LITTLE DIFFICULTY IN BEING ACCEPTED AT A COLLEGE OF YOUR CHOICE.

YOU EARNED A SCORE OF 68 ON THE VERBAL SECTION OF THE TEST. THE VERBAL SECTION OF THE TEST MEASURES YOUR ABILITY TO READ WITH UNDERSTANDING AND TO USE WORDS EFFECTIVELY. A SCORE AS HIGH OR HIGHER THAN THE ONE YOU HAVE EARNED IS EARNED BY LESS THAN 11 PER CENT OF JUNIORS OF YOUR SEX WHO LATER ENTER COLLEGE. VERBAL APTITUDE IS PARTICULARLY IMPORTANT FOR SUCCESSFUL COLLEGE WORK IN THE HUMANITIES AND FINE ARTS.

YOU EARNED A SCORE OF 68 ON THE MATHEMATICAL SECTION OF THE TEST. THE MATHEMATICAL SECTION OF THE TEST MEASURES YOUR ABILITY TO REASON AND WORK EFFECTIVELY WITH NUMBERS. A SCORE AS HIGH OR HIGHER THAN THE ONE YOU HAVE EARNED IS EARNED BY LESS THAN 11 PER CENT OF JUNIORS OF YOUR SEX WHO LATER ENTER COLLEGE. MATHEMATICAL APTITUDE IS PARTICULARLY IMPORTANT FOR SUCCESSFUL COLLEGE WORK IN THE SCIENCES AND ENGINEERING.

YOU SHOULD NOT THINK OF YOUR TEST AS EXACT POINTS BUT AS A RANGE OF SCORES EXTENDING ABOUT THREE POINTS ABOVE AND THREE POINTS BELOW THE SCORE WE HAVE REPORTED TO YOU. THEY GIVE A GOOD INDICATION OF HOW YOU MAY EXPECT TO SCORE ON THE SCHOLASTIC APTITUDE TEST. THE CHANCES ARE FOUR OUT OF FIVE THAT YOU WILL SCORE BETWEEN 630 AND 730 ON THE VERBAL SECTION OF THE SCHOLASTIC APTITUDE TEST WHEN YOU TAKE IT NEXT YEAR. THE CHANCES ARE FOUR OUT OF FIVE THAT YOU WILL SCORE BETWEEN 630 AND 730 ON THE MATHEMATICAL SECTION OF THE SCHOLASTIC APTITUDE TEST WHEN YOU TAKE IT NEXT YEAR. INSOFAR AS VERBAL AND MATHEMATICAL APTITUDES ARE CONCERNED YOU CAN HAVE CONFIDENCE IN YOUR ABILITY TO DO SUCCESSFUL COLLEGE WORK.

THE SCORES YOU HAVE EARNED SHOULD ENCOURAGE YOU TO APPLY FOR ADMISSION TO AN OUTSTANDING COLLEGE. YOUR FUTURE EDUCATIONAL PLANS SHOULD CONSIDER ADVANCED GRADUATE WORK. AFTER YOU HAVE DISCUSSED YOUR SCORES WITH YOUR PARENTS AND YOUR COUNSELOR OR PRINCIPAL SHOULD YOU HAVE ADDITIONAL QUESTIONS ABOUT THEM YOU MAY WRITE TO EDUCATIONAL TESTING SERVICE, PRINCETON, N. J.

SINCERELY YOURS,

E. T. S.

Programmed Composition of Psychological Test Reports

A computer-generated verbal diagnostic report on a standardized psychological test has been used routinely at the Mayo Clinic in Rochester, Minnesota since 1962 (Swenson, 1962), and is available commercially from Behaviordyne, Inc. The Mayo program utilizes the Minnesota Multi-phasic Personality Inventory (MMPI) which is an objective pencil-paper psychological test. The machine-produced report is a group of disconnected statements, or decisions, about the subject as measured by the scales of the MMPI. Joseph C. Finney (1966), who is associated with the Department of Psychiatry of the University of Kentucky, improved the Mayo program by offering an alternative of adding scales from the California Psychological Inventory (CPI), and improving the coherence of the report. A large number of scales are scored, 101 with the MMPI alone or 124 with the MMPI and the CPI.

The report is built by selecting statements and then combining them into paragraphs. For each of the 101 scales, one statement is chosen from among eight possible statements, depending on the individual's score on the scale. By this method, 101 statements are chosen from a repertory of 808. Finney's program has the computer compose a full report on each individual's personality--the kind of report that a psychologist might write after seeing a person several times and administering a full battery of tests. The first one-third of a sample report is given as Figure II.3.

Finney and his co-worker, Auvenshine, have subsequently developed several different kinds of reports written for different purposes. Finney and Auvenshine are now extending their work to other objective psychological tests.

Baker (1971) suggested that electronic computer technology be utilized to generate reports on other types of testing instruments in order to make the results more meaningful to the persons examined and to facilitate better use of the results of testing. He concluded:

> The mechanics of having the computer program prepare verbal descriptions depends upon several factors. First, the insight of the test constructor into the area of interest; second, the relation of levels of test and diagnostic scores to pupil performance; third, the cleverness of the computer programmer in generating connecting prose from somewhat disconnected verbal descriptions.

Figure II.3

SAMPLE COPY OF FINNEY'S REPORT ON THE MINNESOTA MULTI-PHASIC PERSONALITY INVENTORY

This is a report of MMPI and CPI testing of a female age 16, case number B-000005. This test, like any test, is subject to error. Testing only supplements other diagnostic examinations.

First let us examine the evidence of validity and the attitude with which she took the test.

RA-0, Cm-4
Gbs-10
L-3
Gi-4
Wb-3

On the CPI she gives mostly the common and conventional answers. That may be a sign of at least average common sense and judgment, and of being sufficiently steady, reliable, and realistic. She does not give a consistently favorable nor a consistently unfavorable picture of herself. She is afraid to admit even small flaws in herself, in terms of standards which are naive, rigid, perfectionistic, moralistic, unrealistic, and overly conventional. That shows a lack of insight. It also shows that she does not distinguish clearly between fundamental obligations, which people can and will expect her to meet, and the lesser or shallower matters in which a falling short in performance is tolerable. She wants to make a good impression in taking the test, and she gives the impression of having at least an average degree of warmth. She is moderately ambitious, alert, and productive, and likes working. She has no serious doubts about herself.

A-5, 7-4[c]
R-4, Rhw-5[c]
B-4[c], Cfi-5
3f-27

She does not tell of anxiety or stress and is not looking for help. She is a reasonably compliant person. She has a normal amount of flexibility.

In terms of these factors, she seems to be a normal, average, flexible person.

2pc-31

But the two-point code tells us as follows. She has hysterical conversion reactions of some specific location or other. She is naive, exhibitionistic, self-centered, and demanding, and tends to manipulate and exploit people. Because of repression she lacks insight and is not motivated for psychotherapy.

Now, what is the evidence for psychosis or mental illness?

Rcc
Scpt-3
Sx-3
43-7

None of the measures indicate that she is psychotic. Some measures are doubtful, as follows. The obsessive and schizophrenic indicators are about at equal level. On a schizophrenic correction scale, she seems to have psychotic trends. But most measures indicate that she is not psychotic. She does not use rituals or compulsive acts at all to ward off anxiety.

Next we consider narcissism, guilt, and basic trust.

N-7
Sa-5
29-6
Gu-6
27-4
Em-7, 4-5[c]
Dom-4, Es-5
6-3

Her self-esteem is low and she doesn't feel proud of herself. But by another measure her self-acceptance is within the average range, though she tends to blame herself a little more than the average. She shows signs of less than average guilt feeling. And she tends to deny guilt. She has a normal amount of concern with what people think of her. But she denies any feeling of self-consciousness or embarrassment. She has at least average dominance and initiative. She scores almost average on ego strength, and has fair tolerance for frustration. This is a good level of ego strength for a psychiatric patient. She has the assets of benefit from psychotherapy, but only if motivation and distress are also present. To a moderate degree she maintains an optimistic attitude by denying discouragement. She tells of very little worrying; less than the average person.

PtK-4[c]
44-4[c], Ob-6
42-3, Ph-6

She shows signs of having some fears or phobias. But she does not admit fears or phobias.

Now we turn our attention to problems of dependency.

1-4
2-4
Rfd-5[c], 3-4[c]
Ul-5[c], Ody-4[c]

The signs are that she has only slightly more dependency need than the average if at all so. Within the average or normal range she seems to put her dependency needs into action.

Now, what about being demanding or orally aggressive?

The Madison Project

The purpose of the Madison Project (Mathews 1971, 1972) was to design, produce, and evaluate testing reports that would be useful and meaningful at the classroom level: reports that would be useful to teachers and informative to parents, without making time demands on teachers in either preparation of the reports or inservice training. Testing reports were designed for two audiences; teachers and parents. In both cases, these reports were in a narrative format that was uniquely generated by a computer on the basis of individual pupil scores on the fourth-grade level of the Iowa Tests of Basic Skills. A sample of the parent report is given in Figure II.4.

Verbal descriptors were chosen to correspond with the associated ranges of local percentiles. The descriptors used were "excellent," "very strong," "quite strong," "near average," "rather weak," and "very weak."

A basic form for the explanatory paragraphs was developed, and then several equivalents of each paragraph were written to provide variety of style.

The lead paragraph reported the total score the student earned in terms of local and national percentiles. The function of the second paragraph was to describe the pupil's performance in the five main areas of the test: vocabulary, reading comprehension, language skills, work-study skills, and arithmetic. The third paragraph characterized the pupil's attainment in the language skills tests of spelling, capitalization, and word usage. The next paragraph depicted the student's achievement in three specific work-study skills measured by the test battery: map reading, reading graphs and tables, and knowledge and use of reference materials. The fifth paragraph reported the pupil's performance in arithmetic concepts and problem solving, and also on the modern mathematics supplementary test.

Following these five narrative paragraphs, two lists of diagnostic areas were printed. Diagnostic areas were determined by item analyzing the test and setting criterion levels of performance for each of the 25 diagnostic skills. Depending upon a pupil's performance on the test items in each diagnostic area, he was rated as proficient or deficient. For diagnostic areas in which the student showed average performance, no rating was given.

The report for the teacher was very similar to the report to the parents, with the exception that it also contained percentile ranks for each test area.

Figure II.4

PARENT REPORT ON INDIVIDUAL PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
ADMINISTERED SEPTEMBER 1970
MADISON PUBLIC SCHOOLS
MADISON, WISCONSIN

STUDENT: DANIEL MARSHALL
TEACHER: MRS. ANASTASI

ACCORDING TO THIS TEST, DANIEL'S OVERALL ACHIEVEMENT LEVEL IS AT THE 75 PERCENTILE WHEN COMPARED WITH STUDENTS IN GRADE 4 IN THE MADISON SCHOOL SYSTEM, AND AT THE 79 PERCENTILE WHEN COMPARED WITH A NATIONAL SAMPLE OF GRADE 4 STUDENTS.

THE TEST PROVIDES INFORMATION IN FIVE MAIN AREAS. DANIEL'S VOCABULARY SCORE IS NEAR AVERAGE. IN THE AREA OF LANGUAGE SKILLS HIS SCORE IS QUITE STRONG, AND HIS SCORE IN WORK-STUDY SKILLS IS VERY STRONG. HIS ARITHMETIC SCORE IS EXCEPTIONAL.

THE TEST LOOKED AT FOUR SPECIFIC LANGUAGE SKILLS, AND THIS STUDENT WAS RATED AS FOLLOWS: QUITE STRONG IN SPELLING AND QUITE STRONG IN CAPITALIZATION. RATHER WEAK IN PUNCTUATION AND NEAR AVERAGE IN WORD USAGE.

IN THE AREA OF WORK-STUDY SKILLS, THREE SPECIFIC SKILLS WERE TESTED. IN MAP READING, DANIEL SEEMS TO BE QUITE STRONG WHILE HE IS NEAR AVERAGE IN READING GRAPHS AND TABLES. IN HIS KNOWLEDGE AND USE OF REFERENCE MATERIALS HE IS QUITE STRONG.

IN THE AREA OF ARITHMETIC, HE APPEARS TO BE NEAR AVERAGE IN CONCEPTS AND VERY STRONG IN PROBLEM SOLVING WHILE HE TENDS TO BE EXCEPTIONAL IN THE MODERN MATHEMATICS SUPPLEMENTARY TEST.

DIAGNOSTIC KEYS INDICATE THAT DANIEL IS PROFICIENT IN:

SOLVING PROBLEMS INVOLVING MONEY.
SUBTRACTION.
ADDITION.
UNDERSTANDING THE CONCEPT OF WHOLE NUMBER IN MODERN MATH.

DIAGNOSTICS ALSO INDICATE THAT DANIEL IS DEFICIENT IN:

SPELLING BECAUSE OF PROBLEMS IN VOWEL SUBSTITUTION.
USING THE COMMA.
PUNCTUATION BECAUSE OF OVERPUNCTUATION.

STUDENT PERFORMANCE IS EVALUATED IN SEVERAL WAYS: BY THE TEACHER, BY THE PARENTS OF THE STUDENT AND, OF COURSE, BY THE STUDENT HIMSELF. TEST RESULTS DO NOT REPLACE ANY OF THESE, BUT CAN BE VALUABLE IN THAT THEY ARE ANOTHER SOURCE OF DATA. THIS TEST, LIKE ANY TEST, IS SUBJECT TO ERROR.

COMMENTS ABOUT DANIEL MARSHALL FROM MRS. ANASTASI:

One other report was generated for the teacher. This report began with five paragraphs of text similar in design to those already described, but the narrative described the performance of the class as a whole. A sample of this report is given as Figure II.5. This class summary report to the teacher tallied the proficiency and deficiency ratings and also included suggestions that the teacher might use to improve performance of his class in the areas in which their performance was weakest.

The evaluation of the Madison Project concluded that:

1. Each of the experimental testing reports on the Iowa Tests of Basic Skills was preferred by fourth grade teachers in the Madison Public Schools, when compared to the traditional testing reports.

2. Teachers who received the experimental reports reported that they felt better prepared to communicate information obtained from the tests to the parents of their pupils than did teachers who received the traditional reports.

3. Teachers who received the experimental reports showed a greater change in their perception of the performance of their pupils than did the teachers who received the traditional testing reports.

Four projects have been briefly described, and samples of their reports have been shown. In each of them the product was a computer-generated report on testing that was written in a narrative format. The desired advantage of the automated reports was improved communication. These reports have all been found to be effective and efficient vehicles of communication when compared to past practices, but they have only begun to fill the need.

The Madison Project did a respectable job at what it did, but it did not do enough. It made a major step in removing the obstacle of terminology and numerical ratings, but it only reported the pupil's performance compared to other fourth garders. No reference was made to the child's performance compared to his own previous performance. The Madison reports generated the same basic format for all children, making no allowances for different needs at different levels of performance. The greatest shortcoming of these reports, however, was the absence of a pupil report.

Figure II.5

REPORT ON CLASS PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
ADMINISTERED SEPTEMBER 1970
MADISON PUBLIC SCHOOLS

TEACHER: MRS. ANASTASI GRADE 4 SCHOOL: JAMES

THE MEAN ACHIEVEMENT LEVEL OF THE STUDENTS IN YOUR CLASS IS AT THE 54 PERCENTILE WHEN THEIR TEST SCORES ARE COMPARED TO THE OTHER FOURTH-GRADE STUDENTS IN THE MADISON PUBLIC SCHOOL SYSTEM. THE MEAN ACHIEVEMENT LEVEL OF YOUR STUDENTS IS AT THE 55 PERCENTILE WHEN COMPARED WITH A NATIONAL SAMPLE OF FOURTH-GRADE STUDENTS.

THIS TEST PROVIDES INFORMATION IN FIVE MAIN AREAS: IN VOCABULARY, THE MEAN OF THIS CLASS IS NEAR AVERAGE (56 %ILE), WHILE IT IS NEAR AVERAGE IN READING COMPREHENSION (47 %ILE). THE CLASS MEAN IS NEAR AVERAGE IN LANGUAGE SKILLS (55 %ILE) AND NEAR AVERAGE IN WORK-STUDY SKILLS (53 %ILE). THE MEAN SCORE OF THIS GROUP IS NEAR AVERAGE IN ARITHMETIC SKILLS (50 %ILE).

LOOKING A LITTLE CLOSER AT THIS CLASS, LANGUAGE SKILLS, THEY TEND TO BE NEAR AVERAGE IN SPELLING (57 %ILE) AND QUITE STRONG IN CAPITALIZATION (70 %ILE). THERE ARE INDICATIONS THAT THIS CLASS IS NEAR AVERAGE IN PUNCTUATION (57 %ILE) AND NEAR AVERAGE IN WORD USAGE (50 %ILE).

IN THE AREA OF WORK-STUDY SKILLS, THIS GROUP, AS A WHOLE, APPEARS TO BE NEAR AVERAGE IN MAP READING (54 %ILE), NEAR AVERAGE IN READING GRAPHS AND TABLES (55 %ILE), AND NEAR AVERAGE IN KNOWLEDGE AND USE OF REFERENCE MATERIALS (55 %ILE).

WITH REGARD TO THE ARITHMETIC SKILLS OF YOUR STUDENTS, THE CLASS MEAN SHOWS THEM TO BE NEAR AVERAGE IN ARITHMETIC CONCEPTS (55 %ILE) WHILE THEY APPEAR TO BE NEAR AVERAGE IN PROBLEM SOLVING (50 %ILE). ON THE MODERN MATHEMATICS SUPPLEMENT THE SCORES OF THE STUDENTS SEEM TO BE NEAR AVERAGE (57 %ILE).

FOLLOWING ARE THE NUMBERS OF STUDENTS IN YOUR CLASS OF 25 WHO WERE RATED EITHER PROFICIENT OR DEFICIENT IN THE VARIOUS DIAGNOSTIC AREAS:

PROF.	DEF.	DIAGNOSTIC AREA
1	2	RECOGNIZING AND UNDERSTANDING IMPORTANT FACTS AND DETAILS.
2	2	RECOGNIZING AND UNDERSTANDING IMPLIED FACTS AND RELATIONSHIPS.
*	4	SPELLING BECAUSE OF PROBLEMS IN VOWEL SUBSTITUTION.*
2	9	USING VERBS.
5	7	USING VERB FORMS.
4	10	USING THE COMMA.
5	8	USING THE PERIOD.
2	10	USING INTERNAL PUNCTUATION IN A SENTENCE.
3	7	USING TERMINAL PUNCTUATION IN A SENTENCE.
4	8	USING LETTER-WRITING PUNCTUATION.
*	5	USING TOO MUCH PUNCTUATION.*
16	1	USING CAPITALIZATION IN LETTER WRITING.
10	5	CAPITALIZATION OF GEOGRAPHIC TERMS.
11	3	CAPITALIZATION OF HUMAN NAMES AND REFERENCES TO HUMANS.
0	4	LOCATING PLACES ON MAPS AND GLOBES.
11	2	ALPHABETIZING.
3	6	USING AN INDEX.
0	16	USING A DICTIONARY.

Figure II.5 (con't.)

PROF.	DEF.	
3	12	USING GENERAL REFERENCE MATERIALS.
4	6	UNDERSTANDING THE CONCEPT OF WHOLE NUMBER.
1	8	WORKING PROBLEMS USING WHOLE NUMBERS.
6	7	SOLVING PROBLEMS INVOLVING MONEY.
2	7	SUBTR..CTION.
7	5	ADDITION.
2	6	UNDERSTANDING THE CONCEPT OF WHOLE NUMBER IN MODERN MATHEMATICS.

(* PROFICIENCY IS NOT MEASURED IN THIS AREA.)

TEST RESULTS NEVER REPLACE TEACHER JUDGMENTS. THEY ARE ONLY A HELP TO THE TEACHER IN FORMING MORE RELIABLE JUDGMENTS. THIS TEST, LIKE ANY TEST, IS SUBJECT TO ERROR.

FOLLOWING ARE PRACTICAL SUGGESTIONS FOR IMPROVING THE ACHIEVEMENT LEVELS OF YOUR STUDENTS IN THE AREAS THAT YOUR CLASS AVERAGE WAS BELOW THE MEAN OF ALL THE FOURTH-GRADE PUPILS IN THE MADISON PUBLIC SCHOOLS: (SUGGESTIONS IN OTHER AREAS ARE AVAILABLE FROM THE PROJECT DIRECTOR.)

*SUGGESTIONS FOR DEVELOPING SKILL IN DISCERNING THE PURPOSE OR MAIN IDEA OF A PARAGRAPH OR SELECTION

(CIRCLE ONE)

I USUALLY DO / I DON'T USUALLY 1. USE OBJECTIVE EXERCISES IN WHICH THE PUPILS IDENTIFY THE TOPIC SENTENCE OF A SINGLE PARAGRAPH. IF NECESSARY BEGIN WITH PARAGRAPHS IN WHICH THE TOPIC SENTENCE IS REASONABLY APPARENT.

I USUALLY DO / I DON'T USUALLY 2. HAVE CHILDREN PROVIDE ORALLY OR IN WRITING THEIR OWN STATEMENTS OF THE TOPIC SENTENCE OF SINGLE PARAGRAPHS.

I USUALLY DO / I DON'T USUALLY 3. PROVIDE MULTIPLE-CHOICE EXERCISES IN WHICH PUPILS IDENTIFY THE MAIN IDEA OR IDEAS OF PARAGRAPHS OR LONGER SELECTIONS IN WHICH THERE ARE NO CLEAR-CUT STATEMENTS OF THE CENTRAL IDEA.

I USUALLY DO / I DON'T USUALLY 4. HAVE PUPILS STATE IN THEIR OWN WORDS THE MAIN IDEA OR IDEAS OF PARAGRAPHS OR LONGER SELECTIONS WHICH DO NOT CONTAIN A DEFINITE TOPIC SENTENCE. AS PUPILS DEVELOP SKILL IN DISCERNING THE MAIN IDEA, THEY SHOULD BE GIVEN EXERCISES IN WHICH THE CENTRAL IDEAS ARE LESS OBVIOUS.

I USUALLY DO / I DON'T USUALLY 5. ASK QUESTIONS SUCH AS 'WHAT WAS THE AUTHOR'S PURPOSE IN WRITING THIS ARTICLE?' OR 'WHAT IS THE ONE MOST IMPORTANT THING TO REMEMBER FROM THE SELECTION?'

PLEASE FREELY CONSULT YOUR BUILDING READING RESOURCE TEACHER AND MATHEMATICS CADRE MEMBER ABOUT FURTHER DIAGNOSTIC TESTS THAT ARE AVAILABLE, AND ABOUT SPECIFIC MATERIALS THAT CAN BE MADE AVAILABLE TO YOU FOR USE IN YOUR CLASSROOM. ALSO AVAILABLE TO GIVE YOU SUPPORT IN THESE AREAS ARE:

YOUR REMEDIAL READING CONSULTANT: MRS. ALICE BENN,
YOUR DEVELOPMENTAL READING CONSULTANT: MRS. KATHLEEN HARTY,
YOUR ENGLISH-LANGUAGE ARTS COORDINATOR: MR. LEE HANSEN,
YOUR MATHEMATICS COORDINATOR: MR. PETER CHRISTIANSEN,
YOUR SOCIAL STUDIES COORDINATOR: MR. OMAR KUSSOW,

THEY MAY BE CONTACTED AT THE ADMINISTRATION BUILDING BY CALLING 257-9561.

IF YOU HAVE ANY QUESTIONS OR COMMENTS ABOUT THESE REPORTS, PLEASE CALL MR. MATHEWS, THE PROJECT DIRECTOR, AT 257-9561 EXTENSION 252. WE HOPE THESE REPORTS ARE HELPFUL TO YOU.

REFERENCES

Baker, Frank B., "Automation of Test Scoring, Reporting and, Analysis." Chapter 8 in Educational Measurement, 2nd ed., R. L. Thorndike ed. (Washington, D. C.: American Council on Education, 1971).

Finney, Joseph C., "A Programmed Interpretation of the MMPI and the CPI," Archives of General Psychiatry, XV (1966), pp. 202-36.

Helm, Carl E. and Harasymiw, Stefan J., "Computer-based Score Reports," Measurement and Evaluation in Guidance, I, 1 (Spring, 1968), pp. 27-35.

Holmen, Milton G. and Docter, Richard R., "The Testing Industry and Public Policy," a preliminary report published by the Research Institute for Business and Economics, University of Southern California, (October 30, 1968), p. 14.

Kelley, Allen C., "An Experiment with TIPS: A Computer-aided Instructional System for Undergraduate Education," The American Economic Review: Papers and Proceedings, LXXX (1968), pp. 446-57.

Mathews, Walter M., Automated Verbal Testing Reports, (Madison, Wis.: Madison Public Schools, 1971).

Mathews, Walter M., "The Development of Computer-generated Verbal-format Testing Reports," Journal of Educational Data Processing, IX, 3 (1972), pp. 1-11.

Swenson, W. M., et al., "Symposium on Automation Techniques in Personality Assessment," Proceedings of Staff Meetings of the Mayo Clinic, XXXVII (1962), pp. 61-82.

III

THE IOWA PROJECT: IOWA TESTS OF BASIC SKILLS NARRATIVE REPORTS FOR PUPILS

Ralph Balfour

Description of Differential Reports Based on Performance

The success or failure of this project may very well depend upon the degree of appropriate feedback from the system to the pupil. Taking into account the wide range of development among students in the skills of receiving information we found it initially important to develop a program that would, hopefully, provide for these differences, so that we could be reasonably assured that what was intended to be transmitted to the student was received. We assumed that pupils demonstrating different levels of skill development have differing needs for evaluative information and, accordingly, we differentiated three report types, each designed to meet the perceived needs of pupils in a performance range level. "Normal curve" derivations were used to divide our population into three major groups on the composite score scale of the Iowa Tests of Basic Skills. These groups are called, for the time at least, the high, the middle, and the low ranges, and they correspond to the 99th through 59th percentiles, the 69th through the 29th percentiles, and the 39th through the 1st percentiles respectively. Each range, therefore, yields a 40 percentile band. Because of the possibilities of errors of measurement and other technical concerns, which this report will not expand upon, we felt that a 10 percentile overlap between the middle and high, and middle and low ranges would provide flexibility by allowing the teacher to decide which type of report would be most appropriate for a given student. Those pupils are identified by subscripts corresponding to $Pupil_1$ = low, $Pupil_2$ = middle, $Pupil_3$ = high. Samples of the three report types are presented as Figures III.1-III.3.

Two Conceptual Models Employed to Generate Test Feedback: Performance, Growth

Performance statements are based on a student's percentile rank in relation to other students in the distribution of scores in a given skill area. Performance is reported in terms of verbal descriptors corresponding to percentile bands, based on Iowa norms. These are:

Verbal Descriptor	Percentile Band
Exceptional	90-99
Very Strong	80-89
Strong	65-79
Near Average	35-64
Weak	15-34
Very Weak	0-14

Figure III.1

PUPIL REPORT ON PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
Administered January 1973
Iowa City Public Schools
Iowa City, Iowa

HORACE MANN SCHOOL

Pupil: JOHN BLACK GRADE 5 Teacher: MRS. RICHARDSON

A few weeks ago, you took the Iowa Tests of Basic Skills. You may remember that it had questions on vocabulary, spelling, capitalization, punctuation, and reading maps, graphs, and tables. There were also questions about short stories you were asked to read, and there were problems in arithmetic, too.

These tests give you a chance to find out how you are doing in the skills that are used in school. Your teacher gave the tests to everyone in the class to help her see where the class needs most work. Another reason for giving the tests was to give you information on your skills--which skills are strong, and which ones are weak. It's good to find out which skills need attention, because then you can work to improve them.

This report will tell you how you did in the tests and how much you have grown in each area since last year. Your score is compared to the score you had on a similar test you took in the fourth grade.

VOCABULARY

This test measures your understanding of the meaning of words. Your growth during the past year was below average. One good way to improve your vocabulary would be to take a little time to look in the dictionary for the meaning of words you are not sure of. It takes extra time, and sometimes you might have to look up the same word several times before you remember its meaning. But your vocabulary will get bigger and you will better understand what you read and hear. And you will be better able to express what you think.

READING COMPREHENSION

This test measures your ability to understand ideas in what you read. Your growth in this area was below average. One way to improve in this area is to read out loud to a friend, and compare your understanding of what you read with the understanding he or she has.

LANGUAGE SKILLS

Four tests measured how well you understand the tools of writing. These were tests in Spelling, Capitalization, Punctuation, and Word Usage. Your overall growth in these skills was average. Let's look at how you did in each skill.

SPELLING--Your growth in this skill was below average. If you really want to improve your spelling you can! When you are not sure how to spell a word, look it up in the dictionary. Soon you will find you will not have to look up words as often.

Figure III.1 (Con't.)

CAPITALIZATION--Your growth in this skill was average. Keep up the good work.

PUNCTUATION--Since last year, your growth in using punctuation was average. That means you have made good progress in this area.

WORD USAGE--This test measures your use of pronouns, verbs, adjectives and adverbs. Your growth in this area was below average since last year.

WORK-STUDY SKILLS

There were three tests that measure work-study skills: a test on reading maps, a test on reading graphs and tables, and a test of your knowledge and use of reference materials. Your growth in these skills since last year was average. Let's look at how you did in each test:

MAP READING--In this test your growth was average. Keep up the good work.

READING GRAPHS AND TABLES--In this skill your growth was above average. You can feel proud of this improvement.

REFERENCE MATERIALS--This test measures your understanding of how to find information. You showed average growth in this area. That's good progress.

MATHEMATICS SKILLS

Your overall growth in the two tests of arithmetic skills was below average.

MATHEMATICS CONCEPTS--On the test that measures your basic understanding of arithmetic, you have shown average growth since last year. That's good improvement.

MATHEMATICS PROBLEM SOLVING--Your growth in the skill of solving arithmetic problems was below average. Your teacher can help you in this skill in many ways, if you ask.

In general, these tests show that you have done very well in improving your work-study skills, especially reading graphs and tables, but not so well in improving vocabulary and reading skills. You also want to pay attention to spelling and word usage, and also to arithmetic problem solving.

Of all the five test areas, you did best in work-study skills, especially reading graphs and tables. You need to give more attention to your vocabulary, reading and math skills.

There are other things you do at school that are also important. You are growing in many ways, not only in your skills, but also as a person. The Iowa Tests of Basic Skills do not measure these other areas of growth, so they have not been talked about here. It's important that you, your parents, and your teacher talk about these other areas--your everyday activities in all areas of school, your interests, and how you feel about school. That way, you can be more sure that you are getting the most out of your experience in school.

Figure III.2

PUPIL REPORT ON PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
2
Administered January 1973
Iowa City Public Schools
Iowa City, Iowa

HORACE MANN SCHOOL

Pupil: JOHN BLACK GRADE 5 Teacher: MRS. RICHARDSON

A few weeks ago, you took the Iowa Tests of Basic Skills. You may remember that it had questions on vocabulary, spelling, capitalization, punctuation, and reading maps, graphs, and tables. There were also questions about short stories you were asked to read, and there were problems in arithmetic, too.

These tests give you a chance to find out how you are doing in the skills that are used in school. Your teacher gave the tests to everyone in the class to help her see where the class needs most work. Another reason for giving the tests was to give you information on your skills--which skills are strong, and which ones are weak. It's good to find out which skills need attention, because then you can work to improve them.

This report will tell you how you did in the tests. The first comment for each test tells you how strong or weak your score was in that test. Your score is compared to the scores of other students in sixth grade in Iowa. Your score is also compared to the score you had on a similar test you took in the fourth grade. This shows how much you have grown in that area since last year.

VOCABULARY

In the test on vocabulary, your score was near average. Your growth during the past year was below average. One good way to improve your vocabulary would be to take a little time to look in the dictionary for the meaning of words you are not sure of. It takes extra time, and sometimes you might have to look up the same word several times before you remember its meaning. But your vocabulary will get bigger and you will better understand what you read and hear. And you will be better able to express what you think.

READING COMPREHENSION

Your performance on this test was strong. You showed normal growth in reading in the past year. In this test you did very well in seeing and understanding important facts and details in what you read.

LANGUAGE SKILLS

On the four tests that measured the mechanics of writing, your overall performance was near average. Your growth from last year was below average. These are some areas where a little extra work would pay off.

SPELLING--Your score on this test was very weak. You showed minimum growth since last year. It appears that you sometimes are not

Figure III.2 (Con't.)

sure which vowels correctly spell a word. If you really want to improve your spelling, you can! Use a dictionary to check spellings. Your teacher will give you help too, if you want it.

CAPITALIZATION--In capitalization your performance was near average. You showed normal growth.

PUNCTUATION--Your performance in this area of language skills was less than average. Your growth was below average. You tended to use too much punctuation.

USAGE--Your performance in the proper use of words was near average. Your growth since last year was normal. In this test you did very well in using verbs.

WORK-STUDY SKILLS

There were three tests that measure work-study skills: a map reading test, a test of reading graphs and tables, and a test of your knowledge and use of reference materials. In general, your performance was near average. You showed normal growth.

MAP READING--Your score on this test was near average. Your growth since last year was below average. You seemed to have trouble locating places on maps and globes. Some practice in using a map of Iowa City and the state of Iowa would be helpful. (The attendant at a gas station might give you a map if you asked for it.)

READING GRAPHS AND TABLES--In this test your score was strong. Your growth in this area was above average. You've improved in this area and should be proud of it.

REFERENCE MATERIALS--Your knowledge and use of reference materials is near average, according to the test. You've shown normal growth since last year. You did miss questions on the use of an index.

MATHEMATICS SKILLS

Your overall performance in the two tests of arithmetic skills shows that you are strong in this area. Your growth in these skills since last year was above average. Keep it up!

MATHEMATICS CONCEPTS--Your score on this test was strong. You showed above average growth in this area.

MATHEMATICS PROBLEM SOLVING--You have shown strong performance in this area. Your growth in this skill was above average. You did well in solving problems involving money.

In general, these tests show that you have done very well in improving your mathematics skills in this past year, but not so well in vocabulary and language skills. You will want to pay close attention to spelling and punctuation. You also showed good progress in reading graphs and tables.

Figure III.2 (Con't.)

Of all the five test areas, you did best in mathematics. You need to improve the most in work-study skills, especially map reading and using reference materials. You could also improve your language skills, mainly spelling.

There are other things you do at school that are also important. You are growing in many ways, not only in your skills, but also as a person. The Iowa Tests of Basic Skills do not measure these other areas of growth, so they have not been talked about here. It's important that you, your parents, and your teacher talk about these other areas--your everyday activities in all areas of school, your interests, and how you feel about school. That way, all of you can be more sure that you are getting the most out of your experience in school.

Figure III.3

PUPIL$_3$ REPORT ON PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
Administered January 1973
Iowa City Public Schools
Iowa City, Iowa

HORACE MANN SCHOOL

Pupil: JOHN BLACK GRADE 5 Teacher: MRS. RICHARDSON

A few weeks ago, you took the Iowa Tests of Basic Skills. You may remember that it had questions on vocabulary, spelling, capitalization, punctuation, and reading maps, graphs, and tables. There were also questions about short stories you were asked to read, and there were problems in arithmetic, too.

These tests give you a chance to find out how you are doing in the skills that are used in school. Your teacher gave the tests to everyone in the class to help her see where the class needs most work. Another reason for giving the tests was to give you information on your skills--which skills are strong, and which ones are weak. It is good to find out which skills need attention, because then you can work to improve them.

This report will tell you how you did in the tests. The first comment for each test tells you how strong or weak your score was in that test. Your score is compared to the scores of other students in sixth grade in Iowa. Your score is also compared to the score you had on a similar test you took in the fourth grade. This shows how much you have grown in that area since last year.

VOCABULARY — Your performance in this test was very strong. Your growth was above average.

READING COMPREHENSION — Your performance in this test was strong. Your growth was average. You need to give more attention to recognizing and understanding implied facts and details.

LANGUAGE SKILLS — Your performance in these tests--Spelling, Punctuation, Capitalization, and Word Usage--was very strong. Your growth was above average for all four areas. You need to be careful of over-punctuation.

WORK-STUDY SKILLS — Your performance on this test was strong. Your growth was average. You need work in comparing values read from a graph or table.

MATHEMATICS SKILLS — Your performance on this test was exceptional. Your growth was above average.

In general, these tests show that you have done very well in improving your skills in all of the areas tested. You will want to pay close attention to recognizing and understanding implied facts and details, punctuation, and comparing values read from a graph or table.

Figure III.3 (Con't.)

Of the five main test areas, you did best in mathematics.

There are other things you do at school that are also important. You are growing in many ways, not only in your skills, but also as a person. The Iowa Tests of Basic Skills do not measure these other areas of growth, so they have not been talked about here. It's important that you, your parents, and your teacher talk about these other areas--your everyday activities in all areas of school, your interests, and how you feel about school. That way, you can be more sure that you are getting the most out of your experience in school.

The high and middle range pupils are the only recipients of this feedback. It is suppressed on the low range reports to prevent such pupils from receiving an overwhelming quantity of negative evaluative information. It is felt that the competitive nature of this information can be threatening and thereby dysfunctional for the bottom-range student; that such information is only useful for those who have a working understanding of the skills required to compete, as indicated by the level of performance on these skills in a testing situation. The low-level pupil report simply gives a one-sentence explanation of the general skill tested.

Growth statements are based on a comparison between a student's own present and previous year percentile scores in a given skill area. By looking at the empirical evidence for the amount of growth in each test from one grade to the next, at various performance levels, we hope to classify score changes according to the following scale:

Top 5%	Exceptional
Next 20%	Above average
Middle 50%	Average
Next 20%	Below Average
Bottom 5%	Minimum

For students who have very high previous-year scores, the two top verbal descriptors will be pooled and reported as "very strong." For those with very low previous-year scores, the two bottom descriptors will be pooled and reported as "weak." The nature of the growth statements, then, varies with the level of performance. Growth, which is thus individualistic and thereby non-competitive, is reported to all levels.

Tri-Evaluative Methodological Procedures

Ninteen Diagnostic Areas were isolated for Form 5 of Level 11 (5th grade) of the ITBS. Each is a specific skill area in which a minimum of nine items could be classified. (Figure III.4 lists these diagnostic skill areas.)

Norms for proficiency and deficiency in these skills will be built for different levels of pupil composite performance. Criterion levels for proficiency and deficiency will then be determined by examining frequency distributions of raw scores for each diagnostic category. 15-20% will be classified proficient, and the same percentage deficient. Statements will be generated for a skill area only if the pupil's score on that scale is deficient or proficient for his performance level. However, deficiency in a diagnostic area is reported only to pupils falling in the high and middle performance ranges, while proficiencies are reported to all pupils.

A central concern in this approach is that pupils in the low range may well have a number of diagnostic deficiencies, but that it would be

Figure III.4

List of Diagnostic Skill Areas, ITBS, Level 11, Form 5

READING COMPREHENSION

1. Recognize and understand important facts and details
2. Recognize and understand implied facts and details
3. Learning the purpose or main idea of a selection

LANGUAGE SKILLS

4. Vowel substitutions
5. Use of the period
6. Use of the comma
7. Over-punctuation
8. Use of verbs

WORK-STUDY SKILLS

9. Locating places on maps and globes
10. Comparing values read from a graph or table
11. Alphabetizing
12. Use of a dictionary
13. Use of an index

MATHEMATICS

14. Fractions
15. Numeration and number system
16. Problems involving currency
17. Problems involving whole numbers
18. Problems involving successive operations
19. Problems involving subtraction

counter-productive to overload them with negative feedback. A further consideration here is that these diagnostic areas are not exhaustive of the specific skills being tested, and that we are therefore reluctant to pinpoint diagnostic areas from this list for a pupil whose level of skill development is generally weak, and who most likely is experiencing many basic difficulties.

Suggestions are generated from the system to the student to assist him in better communicating his scholastic needs to his parents and/or teacher, and to point out activities that will give him practice in a skill area. None of these remedial suggestions will be received by the high range student for it is assumed that he is aware of methods of improving skills. The middle range student will receive suggestions only on those skills that score below his composite score, and the low student will receive suggestions on the lowest 4 of his 7 scores.

Comments--positive comments only--are generated differentially on the middle and low level reports. They are generated in response to positive growth statements. Three comment types are used: Type I is a short comment--"Keep up the good work," or "That's good improvement." Type II comments are full statements--"That kind of improvement is important," or "You can feel proud of this improvement." Type III comments are still more emphatic--"You've made great gains in this area and should be proud of it." Comments are withheld altogether on the high level reports. The table below shows how they are used on the middle and low level reports.

Level of Report	Test Areas	Level of Growth	Type of Comments
High	---	---	None
Middle	5 Main	Superior	III
Middle	5 Main	Above Average	II
Low	5 Main	Superior or Above Average	III
Low	9 Subtests	Superior	III
Low	9 Subtests	Above Average	II
Low	9 Subtests	Average	I

Pupil Report Format

In addition to the information on each test area, the actual report received by the pupil will contain the following three feedback modes.

The first is a series of short paragraphs which summarize performance. If any of the five main test area scores are significantly high or low, (α=3), they are reported, respectively, as the pupil's best area, or the one most in need of work. If none are significantly high or low, it is commented that his scores show consistency. A similar comparison is made for the nine subtest scores; those significantly high or

low are reported, except on the low level report where significantly low subtest scores are not mentioned in this summary. The second mode consists of a single paragraph which summarizes unusually great or minimal growth, and the final mode is a paragraph which attempts to put the test results in overall perspective, and to encourage communication with parents and teachers. The concern in this entire system is to give pertinent feedback while maintaining the positive self-image necessary for continued growth and development.

IV

NARRATIVE TEST REPORTS FOR PARENTS

Earl Stephan

The parent report is different from the reports for teacher, pupil, or administrator. Obviously, there are certain presuppositions upon which each of these reports are planned. At the risk of repeating some things which already have been stated, it is important to look at the assumptions underlying the parent report.

With all of the adverse publicity which standardized testing is receiving across the nation, perhaps we would do well to issue no parent reports and hopefully draw less and less attention to the fact that we are still testing in our schools. Our first assumption, however, is that if there's testing, then there should be meaningful reporting and sharing of the information gained. Our second assumption is that testing is an integral part of our school program because it is--or should be--used in many ways to promote better instruction and a better learning situation. I believe some recent position papers published by the A.P.G.A. also accept this assumption. Based on these assumptions, the parent's report is an answer to one great mis-use of testing, and that is the dis-use of test results. If, in fact, testing sheds some light on the learning situation, then we need to stop hiding test results and rather make them available in a way that will enable a testing program to fulfill its purposes.

Another assumption is that a good report to the parent can reduce some of the mystery which surrounds learning. The great teacher who is an artist in the classroom can make teaching look like child's play or sleight-of-hand magic. An observer might well query, "Just what is it that he's doing?" or the more skeptical might wonder, "Is anything really happening? If so, what?" The intangibleness of the learning process, the remoteness of the classroom, the seemingly unobservable quality of acquired knowledge--all add an aura of the mystical to education. However, this mystery is reduced a great deal when the learning is at least partly described and measured in terms of acquired skills handled with specific degrees of proficiency.

This leads to still another assumption. Once the learning process can be described and in some way measured, we are then in a better position to change and improve it--or, for that matter, to know if it needs any improvement.

Given all of the above assumptions, we come to the one which is most related to the parents' report *per se*. We assume that parents are vital members on the educational team and as such need to know what is happening in terms they can understand. Further, we assume that if some weaknesses are indicated and some suggestions made available, most parents would take a positive and helping stance rather than flog their child through school with a sharp tongue.

Now, these assumptions are not as assured as they might be, for we found ourselves driven back to our philosophical principles many times--more frequently than the above might suggest. They are stated here so that by making them explicit, we might answer in advance some of the "Why did you add this?" kinds of questions.

Figure IV.1 shows a sample of a report to parents. A brief comparison with the pupil's report will show them to be similar in most respects. The type of report to both the parent and pupil will in every case be related to the level of the pupil's performance on the test.

You'll notice that this report, like the one to the pupil, contains no numerical scores. While realizing that there are many parents who could utilize information in numerical form, we believed that there were more who could not. On this point, as on many others, we were willing to let the professional teacher be the judge of whether percentile scores would clutter or clarify a parent's understanding, and provide them if desired.

One of the real problems, of course, is the desire of most parents to have above average children. Like the pigs in Animal Farm, we were faced with the existence of the "more equals" and the "less equals," and the problem of how to describe them. I'm not sure if it is a statistical problem or if we are dealing with one of the "seven deadly sins" when even a term such as "average" may be considered a derogatory descriptor by many parents. This sample report contains some examples of the descriptors used, and they have been presented more fully elsewhere.

The report shown here is for a child at about the 70th percentile. If you want an afternoon's entertainment, try to write a comparable report for a child at about the 10th. The comments that deal with "How did he do?" and "How has he grown?" need to be both realistic and charitable. Both pupil and parents need information about the pupil's skills and their development. We were concerned not only that the information be provided, but in a way that does not permit its easy use against the child in those instances where he is not living up to parental expectations. To help prevent such negative spin-offs, as well as to emphasize individual achievement, the student's growth is emphasized. Ideally, as problems of information storage on pupil populations are worked out, the reporting of growth can be based on more than one year's change. It has frequently been observed that following a year of exceptional growth in a specific area, there may be a year of little or no growth--and conversely.

Figure IV.1

PARENT REPORT ON PUPIL PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
Administered January 1973
Iowa City Public Schools
Iowa City, Iowa

HORACE MANN SCHOOL

Pupil: JOHN BLACK GRADE 5 Teacher: MRS. RICHARDSON

John's school experience is evaluated in many ways and by various people: by his teacher; by you; and by John, himself. Each of you may have formed a somewhat different opinion and for different reasons. Test results do not replace these important judgments, but they provide another kind of information to the total picture of how well John is doing.

Last month, John took the Iowa Tests of Basic Skills, a series of tests covering a wide range of skills that are important for his future educational development. This report will give you a description of John's performance as compared to that of other fifth grade students and also as compared to his own performance last year. Remember, however, that the Iowa Tests of Basic Skills, like any test, can only sample some kinds of school performance. No test can tell the whole story.

For each test area of the Iowa Tests of Basic Skills, two kinds of information about John's performance are provided. First, a performance statement--his performance in each skill is compared to that of other fifth graders in Iowa. Second, a growth statement--his performance this year is compared to his performance last year in order to see his growth in each skill.

VOCABULARY — John's performance on this test was near average compared to the performance of other fifth grade pupils in Iowa. He showed below average growth in this area since last year. You might encourage him to look in a dictionary for word meanings whenever he is uncertain.

READING COMPREHENSION — On this test, John's performance was strong. His scores showed normal growth in this area in the past year. He did very well in recognizing and understanding important facts and details in what he read.

LANGUAGE SKILLS — These tests measured John's skills in the mechanics of writing. His overall performance in language skills was near average. His growth from last year was below average. Encourage John to take the time to describe fully an event he has experienced. He may then be encouraged to put it in writing and mail it to a friend or relative.

SPELLING--John's performance in the spelling test was very weak. He showed very little growth since last year. He tends not to notice wrong vowels in misspelled words.

Figure IV.1 (Con't.)

CAPITALIZATION--In this test John performed near average. He showed normal growth in capitalization in the past year.

PUNCTUATION--John performed near average in the test of punctuation. His growth was less than average. He tends to include punctuation when it is not necessary. This is usually an indication that the pupil is not sure of the rules of punctuation.

USAGE--In the test of the proper use of words (for example, nouns, verbs, and adjectives), John performed near average. His growth in this area was normal. He did very well in the use of verbs.

WORK-STUDY SKILLS

On these tests of study methods, his performance was near average. He showed normal growth in this area since last year.

MAP-READING--John's score was near average in map reading. His growth since last year was below average. One area John could use more work in is locating places on maps and globes. Letting him help read the map when you travel would aid him in this area, particularly if he helped to calculate distances and to interpret the symbols used on the map.

GRAPHS-TABLES--Reading graphs and tables is a skill in which John's performance was strong. His growth in this area was above average.

REFERENCE MATERIALS--John's knowledge and use of reference materials is near average. His growth was near average in this area. It would be good if John developed the habit of looking up answers to questions that he has, by using a dictionary, encyclopedia, atlas, and other reference materials. Use of an index is a specific area in which he needs practice.

MATHEMATICS SKILLS

John's performance in this area shows that his mathematics skills are strong. His growth in mathematics since last year was above average, and that's good improvement.

MATHEMATICS CONCEPTS--In understanding the concepts involved in mathematics, John's performance was strong. He has shown above average growth on this test.

MATHEMATICS PROBLEM SOLVING--John's score in solving mathematical problems was strong. He has shown above average growth in problem solving. He did well in solving problems involving money. He did well in subtraction.

Let's summarize his growth during this past year. John Black progressed very well this past year in his mathematics skills, but not so well in vocabulary

Figure IV.1 (Con't.)

and language skills, particularly spelling and punctuation. He showed good progress in reading graphs and tables.

Now let's summarize John's performance on this test this year. It can be said that he did best in mathematics and he needs improvement the most in his work-study skills (especially map reading and using reference materials). He also could improve his language skills (particularly spelling).

This information used in conjunction with observations provided by you, John's teacher, and John should result in better meeting the goal you all share: helping John make the most of his experience in school.

In order to make this report as accurate and up-to-date as possible, Mrs. Richardson has added her comments below.

This suggests a comment which is tangential but important. It is possible to find whole classes which show little growth in any basic skill in a given teacher's class, but which will have gained such tremendous social skills or moral concepts that a community would be impoverished without her. Somehow, if this should be the case, the parent's report could be made to reflect such a situation.

Since we felt that a frustrated parent would rather use his emotion constructively if he knew how, some suggestions are included concerning activities which could be undertaken to help the child grow in specific areas. An example can be noted on the sample report, in the first paragraph under Language Skills.

The final paragraph in the parent's report refers to comments by the teacher. This could, of course, be verbal and it might not have been included at all, but again, we all felt that an evaluation of the computer-generated report needed to be made by the on-the-scene teacher. Information other than that resulting from one test needs to be added, and this can help protect against another mis-use of a test, namely, taking its results too seriously. The whole developing student must be brought into the picture, not just his isolated academic skills.

This has been a once-over-lightly treatment of the parent's report. We could discuss its length and language level, etc. Obviously, each could have been varied; what you see was rather deliberately settled upon as optimally clear, concise and useful. Improvements in any one of these dimensions seem to result in some sacrifice in another.

Finally, I want to make a few comments on how these reports might be used. We have some definite suggestions on use for the administrator and teachers, and our ideas on this will continue to evolve as pilot projects are attempted. We think an ideal system would be for the teacher to distribute the pupil's report to him in a private conference, preceding a like conference involving both pupil and parent. The parent will have received a letter from the school principal explaining the testing program, establishing a conference appointment, and encouraging the parent's participation. A sample of such a letter is presented as Figure IV.2. During the conference, the reports are distributed and discussed; a booklet of sample test items might be available to show the parent how various kinds of skills were measured. Something positive ought to be emphasized about each pupil. Discussions might end by setting comfortable goals for the pupil, if possible generated largely by him.

However, we feel that should all these wonderful suggestions be totally ignored and the report simply and coldly mailed to the parent, it would still be a great improvement over the present methods of test reporting employed in many schools.

Figure IV.2

OFFICE OF THE PRINCIPAL
HORACE MANN SCHOOL
BLANK AND STONE STREETS
IOWA CITY, IOWA

February 2, 1973

Dear Mr. and Mrs. Black:

This past January, John took the Iowa Tests of Basic Skills along with the other students in his class. This series of tests provides a measurement of growth in vocabulary, reading, the mechanics of writing, methods of study, and mathematics. These basic skills are important, not only for daily living experiences, but also for future learning.

Skill development is a continuous process for all children, but the rate of growth differs widely among children of the same age or grade. Some learn rapidly; others, who may be as conscientious, learn more slowly. Skill development within each individual also progresses at varying rates, so that a child may have stronger skills in certain areas.

Test results should be used to supplement, not to replace, the judgment of Mrs. Richardson. Tests are aids to better instruction, but they should be used along with other information that Mrs. Richardson has about John -- including ideas and observations from you and John -- in order to meet his instructional needs.

We have a written report for you on John's performance on the Iowa Tests of Basic Skills. It is intended to be a progress report so that you, John, and Mrs. Richardson can discuss plans for John's future growth in school. We invite you to visit John's classroom (Room 203) at Horace Mann School on Tuesday afternoon, February 12, at 1:30 P.M. School will not be in session that day, so please invite John to come with you because at that time Mrs. Richardson will have a testing report to give to him, in addition to the one she has for you, and you will be able to discuss these results together. This time is reserved for you, so please try to attend. I know she is eager to tell you and John about John's growth and also to have your help in planning for John's continued development.

If I can be of help to you, please feel free to call me. My office telephone number is 553-3354, and like Mrs. Richardson, I'm interested in working with you to make John's school experiences good ones.

Sincerely,

C.J. Sharp
Principal

V

NARRATIVE TEST REPORTS FOR TEACHERS

Stephanie D. Alnot

Skeptics abound, and detractors, as ever, speak louder than believers; but beneath the clamor, a more general and widespread undercurrent of confidence in standardized tests and their results is discernible. Teachers, it would seem, generally hold positive opinions about standardized tests: the more familiar they are with tests, the more positive their opinions. Yet it appears that the extent to which test results are used, in practice, does not correspond to expressed attitudes (Goslin, 1967, p. 138-39). To the already-mentioned factors of lack of training and lack of time, Goslin would add still another: lack of clearly-defined school policies regarding test results, their use and dissemination (1967, p. 139-140). The reporting system being described is an attempt to answer, in some part at least, the needs suggested by these insufficiencies.

Unlike pupils and parents, teachers generally do receive as a matter of routine their pupil's standardized achievement test results. There is little serious argument about the teacher's need for objective information about pupils' skills; it is the uses and interpretations a teacher may or may not make that are sometimes called into question. A system that provides interpretive information to pupils and their parents--on the assumption that the teacher hasn't the time or training to do so--must take into account the teacher's needs, not only for information, but for information in the same interpretive framework provided for the other reicpients. The teacher's needs in this respect are, of course, somewhat different. He is familiar with the structure of the tests, and the skills tested. And he must review the results for all the pupils in his class. The reports he receives must provide useful information about each pupil's performance, growth, and specific strengths and weaknesses, and must do so with an eye to the amount of time available to devote to test results. The teacher's reports, then, must be as concise as clarity will allow. And they must function additionally as a summary of the content of the corresponding reports for pupils and parents, in order to be useful as a guide to those reports when distributed in conferences. A sample of a report for a teacher is presented as Figure V.1.

The information contained on the report to the teacher is essentially the same as that presented to the pupil and parents; it differs in only a few respects: The most apparent difference is that along with the verbal descriptors of performance, the percentile rank for each test is reported. This provision is, in part, an acknowledgement

Figure V.1

TEACHER REPORT ON INDIVIDUAL PERFORMANCE ON THE IOWA TESTS OF BASIC SKILLS
Administered January 1973
Iowa City Public Schools
Iowa City, Iowa

HORACE MANN SCHOOL

Pupil: JOHN BLACK GRADE 5 Teacher: MRS. RICHARDSON

The first part of each comment refers to John's test performance this year. (The percentiles show his performance compared to other pupils in Iowa.) The growth statements refer to John's test performance this year compared to last year's testing.

VOCABULARY	Near average performance (61%ile) and below average growth.
READING COMPREHENSION	Strong (71%ile) with normal growth. John did very well in recognizing and understanding important facts and details in reading.
LANGUAGE SKILLS	Near average performance (45%ile) with below average growth.
Spelling	Weak (28%ile) with minimal growth. Vowel substitution is an area in which John needs help.
Capitalization	Near average performance (43%ile) with normal growth.
Punctuation	Near average performance (40%ile) with less than average growth. John tends to use too much punctuation.
Usage	Near average performance (62%ile) with normal growth. He did very well in using verbs.
WORK-STUDY SKILLS	Near average performance (44%ile) with normal growth.
Map Reading	Near average performance (34%ile) with below average growth. Locating places on maps and globes in a problem for John.
Graphs-Tables	Strong (73%ile) with above average growth.
Reference Materials	Near average performance (40%ile) with normal growth. Use of an index is an area in which he needs help.
MATHEMATICS SKILLS	Strong (76%ile) with above average growth.
Mathematics Concepts	Strong (72%ile) with above average growth.
Problem Solving	Strong (79%ile) with growth that was above average. He did well in solving problems involving money. He did well in subtraction.

that some teachers may in fact be in the habit of using these scores; partly it anticipates an overreaction to change itself and the sudden and total withdrawal of all numerical data; and partly this provision is in recognition that at least some teachers experience a need to be in possession of "professional" information--a provision for those who may feel threatened by the prospect of equally-well-informed parents.

A second difference will occur on the lowest of the three levels of reports. The rationale for suppressing performance statements for each test area on the pupil and parent reports in this range, does not apply to the report to the teacher, and performance is reported by test area.

A third difference is simply that the comments and suggestions appropriate to the pupil's and parent's reports, are not generated on the teacher report. Suggestions for enrichment and remediation in diagnostic areas have been generated, however, but because the need for such information will vary greatly from teacher to teacher, and for various pupils, these suggestions are provided separately.

In addition to the reports for each individual pupil, the teacher also receives a report containing class summary information. This report appears as Figure V.2. The first part describes the performance of the class as a whole in each of the 14 test areas; the scale used is a broad version of that used to report individual performance; here, the 1st through 39th percentile is rated "below average;" the 40th to 59th percentile range is "average;" and the 60th through 99th percentile range rates "above average." The mean percentile for the class in each test area is also reported.

Growth for a class is summarized by tallying the numbers of pupils attaining above average, average and below average growth in each test area. These ratings are based on the same scale used to report growth on the individual reports, except that here "above average" encompasses both above average and superior, and "below average" includes the minimum rating as well. This growth summary tallies growth as reported on the individual reports, where, it will be recalled, the determination of a growth rating depends on the pupil's overall level of performance.

The second page of the Class Summary Report lists the diagnostic skill areas for the form and level of the test administered. In the summary, which follows, proficiencies and deficiencies are indicated for each pupil, and tallied for each skill area. Unlike the reporting of proficiency or deficiency on the individual reports, the same criterion level is used for all pupils here, regardless of general performance level. Pupils are listed in order of decreasing composite score, so that pupils at the top of the list will tend to show more proficiencies, while those toward the bottom will tend to have more deficiencies indicated. This format facilitates discernment of needs for curricular adjustment, or of patterns for purposes of grouping in subject areas. The teacher will receive in addition a list of the item numbers in each diagnostic skill classification.

Figure V.2

CLASS SUMMARY REPORT ON THE IOWA TESTS OF BASIC SKILLS
HORACE MANN SCHOOL
IOWA CITY PUBLIC SCHOOLS
IOWA CITY, IOWA

Grade 5 January, 1973 Mrs. Richardson

The summary on the first part of the table below rates the performance of your class as a whole in each test area according to the scale: Below average (01-39 %ile), average (40-59 %ile), above average (60-99 %ile). The second part of the table shows the number of your students that had each of the three levels of growth.

	PERFORMANCE		GROWTH (number of students)		
	rating	mean %ile	below average	normal	above average
VOCABULARY	average	47	5	10	5
READING COMPREHENSION	average	52	5	13	2
LANGUAGE SKILLS	average	46	6	12	2
Spelling	below average	38	10	8	2
Capitalization	average	55	8	8	4
Punctuation	average	49	5	15	0
Usage	average	51	2	12	6
WORK-STUDY SKILLS	average	54	3	6	11
Map Reading	average	45	5	12	3
Graphs-Tables	average	58	1	9	10
Reference Materials	average	50	6	8	6
MATHEMATICS SKILLS	above average	61	3	11	6
Mathematics Concepts	average	59	5	12	3
Problem Solving	above average	62	0	14	6

Figure V.2 (Con't.)

List of Diagnostic Skill Areas (Grade 5, Form 5)

READING COMPREHENSION
1. Recognize and understand important facts and details
2. Recognize and understand implied facts and details
3. Learning the purpose or main idea of a selection

LANGUAGE SKILLS
4. Vowel substitutions
5. Use of the period
6. Use of the comma
7. Over punctuation
8. Use of verbs

WORK-STUDY SKILLS
9. Locating places on maps and globes
10. Comparing values read from a graph or table
11. Alphabetizing
12. Use of a dictionary
13. Use of an index

MATHEMATICS
14. Fractions
15. Numeration and number system
16. Problems involving currency
17. Problems involving whole numbers
18. Problems involving successive operations
19. Problems involving subtraction

The following tallies of proficiency and deficiency for the diagnostic areas are determined differently than they were on the individual pupil reports. Here the same criterion levels are used for ALL pupils, and individual pupil performance level is not considered. Therefore, high performing pupils will tend to have more proficiencies listed here, while low performing pupils will have more deficiencies. The pupils are listed below in order of decreasing overall performance on the test battery.

Figure V.2 (Con't.)

	DIAGNOSTIC SKILL AREAS																	
	READING			LANGUAGE SKILLS					WORK-STUDY SKILLS					MATHEMATIC SKILLS					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Jose Poole	+	+	+		+	+			+		+	+			+	+	+	+	+
Cheryl Pierce	+	+	+					+									-		
Carol Arrango	-			+	+					+				+	+		+	+	
Bob Inez		+				+			+		+						+	+	
Jane Woodarski			+					+					+						-
Carl Jones		+			+			-				+			+	+			-
Mary Thomas	+			-				+				-	+						
Marion Quarry			+			-						-				+		+	
Joseph Palumbo					+						+					+		+	
Bea Kelly		+												-					
John Black	+			-			-	+	-				-			+			+
Harold Lewis		+			+							-					-		
Sandra O'Connor				-								-			-				
William Drake			-				-			+							-		
Jerry Fine					+							-		-					
Samuel Roberts				+					-										
Roberta Edwards			-		+						+	-					-		
Julia Schultz	-		-	-				+						-			-		
Myron Gold	-		-	-			-		-		-			-	-		-		
Sally Green	-		-		-		-		-	-	-	-		-	-	-	-	-	-
TOTALS																			
Proficient	4	6	4	2	7	2	*	5	2	2	5	2	2	1	3	5	3	5	2
Deficient	4	0	5	5	1	1	4	1	4	1	2	7	1	5	3	1	7	1	3

*Proficiency is not defined in this area

Lists of suggestions for remediation in each of the eleven tested areas are available. They are generated for the most important three areas having a mean percentile significantly lower than that of the highest area (α=.3). If there are none, then suggestions for two of the lowest five areas are generated. The order of preference, or importance, for this purpose is Reading Comprehension, Mathematics Skills, Vocabulary, Work-Study Skills, and Language Skills.

Considerations guiding the design of reports for the teacher center on providing a maximum of useful information in a format that requires a minimal investment of the teacher's time. Certainly one important determination to be made in a pilot study is the extent to which the reports achieve an acceptable balance between these considerations.

Our underlying assumption has been that information about test results is needed by everyone involved in learning, teaching, and testing. The objective has been the development of a flexible, time-saving, and supportive system that would make use and communication of test results more frequent and more effective. This reporting system represents an attempt to structure not only needed information and interpretation, but a procedure as well.

REFERENCES

Goslin, David A., Teachers and Testing. (New York: Russel Sage Foundation, 1967).

Mathews, Walter M., Automated Verbal Testing Reports. (Madison, Wis.: Madison Public Schools, 1971).

VI

THE FUTURE OF NARRATIVE TEST REPORTS

Thomas B. Taft Jr.

Development of any alternative system in the educational domain needs a thorough tryout prior to its implementation if it is to survive criticism based on its difference. It has been the purpose of the researchers involved in the development of this narrative reporting system to anticipate problems, and to obtain feedback from educators most concerned with the tests on the specific strengths and weaknesses of the narrative reporting format. This report will explicate the proposed trial implementation of the narrative reporting system and the long range possibilities and applications of the narrative reporting format.

Currently-Proposed Modifications (Pre-pilot)

A first step in this explanation is to identify several modifications of the basic narrative system as outlined by the existing model reports. Most of these modifications involve the addition or deletion of certain aspects of the reports in order to increase clarity or decrease the length of the reports.

First, deletion of the Performance and Growth statements on the total Language, Work-Study and Arithmetic Skills would decrease the length of the report and at the same time increase its clarity by emphasizing the specific skills (subtests) covered by these areas. These different skill areas may need differential treatment that could be overlooked if the overall area score is emphasized.

An addition to the report which may facilitate clarity is that of a graphical display of growth. Even though we are trying to get away from the labeling aspect of the Grade Equivalent (GE) score, it seems necessary to provide information in a concise format on the child's growth considered against some developmental criteria. A computer-printed self-adhesive tape with the GE data represented graphically could be attached to a cumulative profile sheet to provide a readily available reference on the student's progress. Both of the above modifications enhance the feedback effectiveness of the verbal reporting technique and therefore improve the role testing can play in the educative process.

The reports as they are currently generated do not include the composite score. The logic behind this is that if the teacher, parent or student were presented with the composite, he might interpret it as being representative of performance on all of the various subtests, while in fact this is frequently not the case, as can be observed by inspecting some actual profiles of students on the ITBS (see Table 1).

It is obvious that students vary both positively and negatively from the composite score. The psychological consequences of providing a set expectation of overall ability have been widely discussed in works by (Holt, 1964), (Silberman, 1970), and (Rosenthal & Jacobson, 1968). Even though the methodology of the latter work is in doubt (Thorndike, 1968), there seems to be little doubt that expectancies do influence behavior in very important ways.

If the composite were to be included, it would have to be presented at the end of the report and in such a manner that it could not be interpreted to mean that the student "was" a certain score level.

Another possible modification is reporting only exceptions to normal growth. Rather than specifically indicating that the student has been growing normally in each area, it could be indicated that unless otherwise reported normal progress should be assumed. Again the purpose of this modification would be to shorten and clarify a report which might be so lengthy that the recipients choose not to read or consider it carefully.

Corresponding with the above concern is the possibility of abbreviating the pupil report in such a manner that one page would encompass the entire report. While this has been suggested as a possible alternative, there are several problems with it. Particularly, we are trying to provide more information to the student by means of a verbal report so that he can use the results constructively in his educational growth. To shorten the report to any great extent would be getting back to what the narrative reporting system hopes to avoid. What probably needs to be done is to involve the student in the evaluation process, so that a narrative report is meaningful enough to maintain his interest. Significant educational feedback could then be facilitated by this test reporting process.

The Pilot

A proposed pilot study to determine the positive and negative aspects of the current system would involve the following: first, to put the preliminary computer programs to a test, so that adaptations and corrections may be planned. The current system involves the generation of differentiated reports, as previously described. Since these reports involve substantial variation, it will be necessary to be sure the proper sequence of computer operations is carried out in the different reports.

TABLE I

Five Typical Score Profiles of Students on the Iowa Tests of Basic Skills

				Language Skills				Work-Study			Arithmetic		Composite
		Vocabulary	Reading	Spelling	Capitalization	Punctuation	Usage	Maps	Graphs	References	Concepts	Problems	
		V	R	L-1	L-2	L-3	L-4	W-1	W-2	W-3	A-1	A-2	
Mary Smith	GE	35	39	57	51	58	53	38	40	38	45	41	42
	PR	52	63	93	86	92	86	64	67	64	84	79	76
Paul Jones	GE	32	23	25	21	22	29	16	22	24	22	18	24
	PR	42	15	20	11	13	36	02	14	15	06	05	14
Carl Peters	GE	38	43	50	61	54	36	45	43	53	34	57	45
	PR	72	73	83	98	88	54	83	77	96	48	99	83
Susan Green	GE	40	36	61	43	60	50	27	40	51	34	41	41
	PR	68	54	97	71	95	81	25	67	94	48	79	72
Jane Moore	GE	58	19	36	24	41	40	51	35	34	45	38	39
	PR	93	06	54	33	66	62	91	51	48	84	65	64

A second goal is to identify the costs involved in the narrative system both in terms of money and time. A concern in this area is to determine the administrative difficulties that exist as a result of the format. For example, if growth statements are to be generated we need previous-year data, which for the transient student will not be available. This can only provide difficulties for the administration.

A third and equally important goal is obtaining teacher, student and parent reaction to a report actually generated by the system. Finally, we need the teacher's evaluation of the group data the system generates. The usefulness and clarity of the diagnostic skills report and the class summary report is important if the teacher is to use them as aids in planning instruction.

The primary method of gathering data will be questionnaires and interviews. The teacher, student and parent populations will be asked to judge the individual reports along two dimensions.

These dimensions are the cognitive, or knowledge, area and also the affective realm. The concern is not only with the enhancement of the individual's knowledge of the content reported by the narrative system and his increased ability to use this knowledge, but additionally with his affective response to the testing procedure as this effects his overall role in the educative process.

A student is concerned with the consequences of his test performance as they influence his immediate and long range goals. The parent is always striving to gain feedback which might aid him in helping his child or at least in understanding the child's position. The teacher needs to know how successfully he is imparting knowledge and attitudes. Knowledge of the student's rank against both a set criterion and his peers is fruitful information for the teacher if he is to be an effective facilitator of growth.

It is also necessary to obtain concise information on the effectiveness of the system for the local administration. Questions must be structured which ascertain the role testing plays, and how it effects their budgetary problems.

Alternative questionnaires need to be developed which assess whether the narrative system is non-threatening, meeting the needs of these various groups.

The sampling involved in a pilot of this system would ideally involve a representative selection of not only different sized school districts, different ethnic groups, etc., but also of the entire testable population. Economically this is unrealistic, however, and instead, the original tryout of the system will most likely be done in Iowa with a small number of school districts representative of various population groups.

While the results will not be totally generalizable, they will certainly allow us to make positive modifications prior to major implementation. The narrative system as designed should be flexible and adaptive, so that changes can be made at later points as a matter of course.

Beyond the Pilot: Implications for Narrative Reporting

Perhaps the preceding discussion, while informative, has been less than ecstatic. George Leonard, in his book Education and Ecstasy, has said that education should be ecstatic, exciting! In this paper, the following considerations come closest to fulfilling that goal.

The real importance of a computer-generated narrative score report is not the report, but rather the concept of providing an alternative and improved communication channel, specifically oriented to facilitate education and personal growth. The implication is if one alternative can be opened, then others should also be possible. Let's now examine some possible alternatives for generating and using narrative reporting.

That different age groups need feedback at various levels of sophistication is well documented in the development literature. It is well within the technological capacities of the computer to generate animated displays on a cathode-ray tube which could communicate test results in a concrete and attention-getting manner to a very young child. He might even be given examples of his performance difficulties on which he could practice. The older student may need other forms of feedback. He may interact with the computer by asking questions about the implications of his performance or he may obtain feedback regarding the content covered and his weaknesses. While a counselor is most qualified to answer such questions, it would be possible to identify certain advice the computer could give to a student who asked certain types of questions, and program this information for use.

Another application would be to integrate aptitude measures commonly used by school systems with information on achievement measures. Aptitude tests frequently have content similar to that found on achievement measures but they are typically taken at different times and for different purposes. A logical step would be to accumulate both aptitude and achievement information and make comparisons between expected performance and actual achievement. While this type of integration could produce an extremely positive source of information, it will be necessary to monitor such a system to make sure that it is not used in a negative manner. The over- and under-achiever would be easily identified and the learning process facilitated by making curriculum content available to the student at an appropriate level for him.

A further application which makes the narrative system an ecstatic educational tool is its possible adaptability to different educative

structures. Individualized instruction could be facilitated by inclusion in the report of an emphasis on areas of content in which a student has not reached a certain criterion performance level. A similar structure, non-graded classrooms, could be handled in a similar manner. The growth aspect of the report could be emphasized almost to the exclusion of norm-group performance statements. The performance statements included could be based on criterion performance in various content areas, as in the individualized approach.

Another important alternative that the narrative system could handle is different reports for different economic-ethnic subpopulations. The current system provides differentiated reports for various performance levels but little attention has been given to altering the language of the reports to allow for subgroup differences. For the feedback to be meaningful and relevant, colloquial phrases developed by a person familiar with the specific environment could be incorporated. This is just another way of making the system responsive and dynamic rather than static.

The last application of narrative reporting to be discussed is the role it can play in vocational testing and decision making. A meaningful verbal report which describes the student's possible career choices and coordinates this with academic achievement information would serve an important and current need for better career decision-making. The system should provide non-judgmental information which would identify realistic choices and also describe the possible processes by which the student could achieve certain goals.

All of the applications described are feasible and useful, but they do not cover the total range of possibilities. Certain philosophical issues are raised by testing in general, and computer-generated narrative reporting is no exception. This report hopefully will serve as a stimulus both for new ideas in the use of narrative reports, and also for philosophical discussion of this form of reporting. While this paper has not tried to answer any of these questions, it acknowledges the need for discussion prior to major implementation of a new system.

REFERENCES

Holt, John, How Children Fail. (New York: Pitman, 1964).

Rosenthal, Robert and Jacobson, Lenore. Pygmalion in the Classroom, (New York: Holt, Rinehart and Winston, 1968).

Silberman, Charles E., Crisis in the Classroom. (New York: Random House, Inc., 1970).

Thorndike, Robert L., "Review of Pygmalion in the Classroom," American Educational Research Journal. (1968), 5, 709-711.